A scalable algorithm for radio-interferometric imaging

Rafael E. Carrillo*, S. Vijay Kartik*, Jean-Philippe Thiran*, and Yves Wiaux†
* Signal Processing Laboratory (LTS5), Ecole polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.
† Institute of Sensors, Signals, and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK.

***Abstract*—In recent works, sparse models and convex optimization techniques have been applied to radio-interferometric (RI) imaging showing the potential to outperform state-of-the-art imaging algorithms in the field. In this paper, we propose a scalable algorithm for RI imaging that offers a highly parallelizable structure paving the way for next-generation high-dimensional data imaging. The proposed algorithm is based on a proximal linear version of the alternating direction method of multipliers (ADMM).**

Next-generation radio telescopes, such as the future Square Kilometer Array (SKA), are designed to achieve much higher dynamic range—and at a higher angular resolution—than current instruments. In addition, these telescopes will acquire a massive amount of data, thus posing challenges for solving large-scale inverse problems in the perspective of image reconstruction. These challenges have provided motivation for vigorous research in the community to reformulate imaging and calibration techniques for radio interferometry (RI) [1].

The RI measurement equation can be discretized as $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{x} \in \mathbb{C}^N$ is the sensed image, $\boldsymbol{y} \in \mathbb{C}^M$ denotes the vector of measured visibilities, $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is a discretization of the measurement operator, which essentially reduces to a Fourier matrix sampled on M spatial continuous frequencies, and $\boldsymbol{n} \in \mathbb{C}^M$ represents the observation noise [2]. The standard image reconstruction algorithm in RI is called CLEAN, which is in essence a variant of the matching pursuit algorithm, and is known to be slow and to provide suboptimal imaging quality. In [3] some of the authors propose an imaging algorithm dubbed sparsity averaging reweighted analysis (SARA) based on average sparsity over multiple bases, showing superior reconstruction qualities relative to state-of-the-art imaging methods in the field. SARA adopts a reweighted ℓ_1 minimization scheme to promote average sparsity. The algorithm solves a sequence of weighted ℓ_1 problems where the weights are essentially the inverse of the values of the solution of the previous problem [3] [4]. The weighted ℓ_1 problem is defined as:

$$\min_{\bar{\boldsymbol{x}}\in\mathbb{R}^N_+} \|\mathsf{W}\boldsymbol{\Psi}^\dagger\bar{\boldsymbol{x}}\|_1 \text{ subject to } \|\boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}}\|_2 \leq \epsilon, \tag{1}$$

where $\boldsymbol{\Psi}^\dagger$ denotes the adjoint operator of the sparsity dictionary $\boldsymbol{\Psi}$, W is a diagonal matrix with positive weights, $\mathbb{R}^N_+$ denotes the positive orthant in $\mathbb{R}^N$ and ϵ is an upper bound on the ℓ_2 norm of the noise, which can be accurately estimated. Hence, we focus our attention on solving problem (1) efficiently, especially for large-scale data problems, *i.e.* when the number of visibilities is large ($M \gg N$).

In [5] we propose an algorithmic framework based on the simultaneous-direction method of multipliers (SDMM) to solve (1). This framework offers a parallel implementation structure that decomposes the original problem into several smaller simple problems, thus allowing implementation in multicore architectures or computing clusters. Although SDMM provides a parallel structure, a major drawback is that it has to solve a linear system at each iteration of the algorithm, hampering the performance for large-scale data problems.

In this work, we propose to use a proximal linear version of the alternating direction method of multipliers (ADMM) [6] to solve (1). ADMM solves the following equivalent problem:

$$\min_{\boldsymbol{x},\boldsymbol{z}} f(\boldsymbol{x}) + h(\boldsymbol{z}) \text{ subject to } \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y}, \tag{2}$$

where $f(\boldsymbol{x}) = \|\mathsf{W}\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1 + i_C(\boldsymbol{x})$, with $C = \mathbb{R}^N_+$, and $h(\boldsymbol{z}) = i_B(\boldsymbol{z})$, with $B = \{\boldsymbol{z} \in \mathbb{R}^M : \|\boldsymbol{z}\|_2 \leq \epsilon\}$. ADMM utilizes the following augmented Lagrangian function:

$$f(\boldsymbol{x}) + h(\boldsymbol{z}) + \frac{1}{\gamma}\langle\boldsymbol{\lambda}, \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{z} - \boldsymbol{y}\rangle + \frac{1}{2\gamma}\|\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{z} - \boldsymbol{y}\|_2^2. \tag{3}$$

At each iteration, the augmented Lagrangian is minimized over $\boldsymbol{x}$ and $\boldsymbol{z}$ separately, one after the other, followed by a dual update for the Lagrange multipliers $\boldsymbol{\lambda}$. Note that the exact minimization on $\boldsymbol{x}$ involves a quadratic term and is thus not straightforward. The proximal linear method solves the minimization problem on $\boldsymbol{x}$ inexactly by linearizing the quadratic penalty term of the augmented Lagrangian at the current point $\boldsymbol{x}^{(t)}$ and adding a proximal term. The iterative update equations for the proximal ADMM are given by

$$\begin{aligned}
\boldsymbol{z}^{(t+1)} &= \text{prox}_{\gamma h}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^{(t)} - \boldsymbol{\lambda}^{(t)}) \\
\boldsymbol{x}^{(t+1)} &= \text{prox}_{\mu\gamma f}(\boldsymbol{x}^{(t)} - \mu\boldsymbol{\Phi}^H(\boldsymbol{\lambda}^{(t)} + \boldsymbol{\Phi}\boldsymbol{x}^{(t)} - \boldsymbol{y} + \boldsymbol{z}^{(t+1)})) \\
\boldsymbol{\lambda}^{(t+1)} &= \boldsymbol{\lambda}^{(t)} + \beta(\boldsymbol{\Phi}\boldsymbol{x}^{(t+1)} - \boldsymbol{y} + \boldsymbol{z}^{(t+1)})
\end{aligned} \tag{4}$$

where prox_h denotes the proximity operator [7] of the function h and β and μ are proper step sizes. For large-scale problems, we propose to split the data vector $\boldsymbol{y}$ and the measurement operator into R blocks in the following manner: $\boldsymbol{y} = [\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_R^T]^T$ and $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1^T, \ldots, \boldsymbol{\Phi}_R^T]^T$, where each $\boldsymbol{y}_i$ is modelled as $\boldsymbol{y}_i = \boldsymbol{\Phi}_i\boldsymbol{x} + \boldsymbol{n}_i$. With this partition the updates for $\boldsymbol{z}$ and $\boldsymbol{\lambda}$ in (4) can be implemented in parallel for R blocks. Also, note that $\boldsymbol{\Phi}^H\boldsymbol{y} = \sum_{i=1}^R \boldsymbol{\Phi}_i^H\boldsymbol{y}_i$, thus the computation of the gradient in the update equation for $\boldsymbol{x}$ in (4) can be done in parallel using a sum reduction approach. These implementations provide both flexibility in memory requirements and a significant gain in terms of speed, thus enabling scalability to large-scale problems.

In the following we present numerical results to illustrate the performance of the proposed algorithm. Figure 1 shows the test image M31 and a sample coverage. Figure 2 shows reconstruction results (SNR and timings) using SARA with the proposed algorithm from simulated visibilities for the test image with 40 dB of input SNR. The results show that the reconstruction SNR improves on the input SNR for $M > 2N$, thus a large number of visibilities are needed to achieve the desired dynamic range in next-generation instruments. Also, it shows that the algorithm scales with M, taking less than 35 minutes to reconstruct an image from a data set of $M = 10N$ visibilities using unoptimized MATLAB code. Preliminary results of the proposed algorithm applied to real data are shown in Figure 3. The reconstructed image is cleaner than dirty image and reveals structures that were hidden.

A possible future step for big data scalability is to incorporate ideas from stochastic gradient methods into ADMM. The key idea is to use only one data block $\boldsymbol{y}_i$, or a subset of blocks, at each iteration of the reconstruction thus reducing the computational complexity. See for example [8] and references therein for first theoretical results.

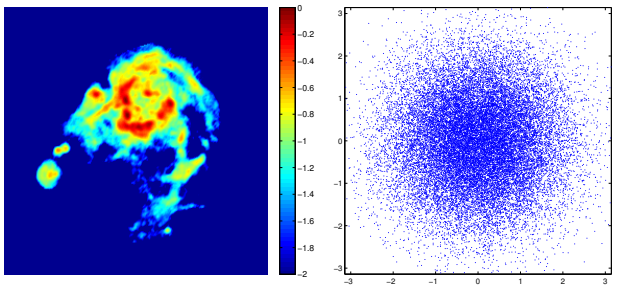

Figure 1. Left: Original 256×256 test image, M31, shown in a $\log_{10}$ scale with normalized peak value. Right: Example of a simulated variable density pattern in the Fourier plane.

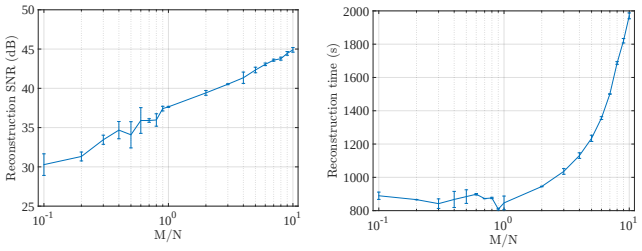

Figure 2. Reconstruction results using simulated measurements ranging from 0.1 to 10 times image size (approximately 6500 to 650000 visibilities). Measurement noise was added with 40 dB of input SNR.

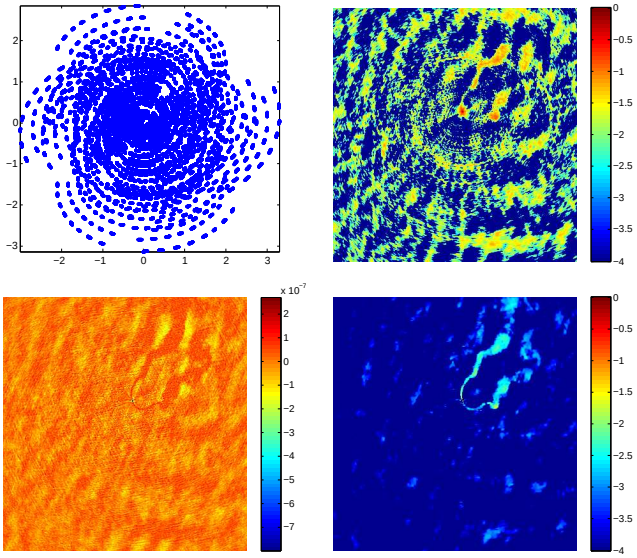

Figure 3. Reconstruction from real telescope data. Top left: sampling pattern in the Fourier plane ($M = 140918$). Top right: dirty image ($\Phi^H y$). Bottom right: reconstructed image. Both images have a 2048×2048 resolution and are zoomed in on the inner part to show the salient structures. Also, both images are shown in a $\log_{10}$ scale with normalized peak value. Bottom left: Residual map ($\Phi^H(y - \Phi \boldsymbol{x})$).

REFERENCES

[1] S. Wijnholds, A.-J. van der Veen, F. De Stefani, E. La Rosa, and A. Farina, "Signal processing challenges for radio astronomical arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 5382–5386.

[2] U. Rau, S. Bhatnagar, M. A. Voronkov, and T. J. Cornwell, "Advances in calibration and imaging techniques in radio interferometry," *Proc. IEEE*, vol. 97, pp. 1472–1481, 2009.

[3] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *MNRAS*, vol. 426, no. 2, pp. 1223–1234, 2012.

[4] R. E. Carrillo, J. D. McEwen, D. V. D. Ville, J.-P. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Process. Letters*, vol. 20, no. 6, pp. 591–594, 2013.

[5] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *MNRAS*, vol. 439, no. 4, pp. 3591–3604, 2014.

[6] J. Yang and Y. Zhang, "Alternating direction algorithms for l1-problems in compressive sensing," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.

[7] P. L. Combettes and J.-C. Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, New York, 2011, ch. Proximal splitting methods in signal processing, pp. 185–212.

[8] S. Azadi and S. Sra, "Towards an optimal stochastic alternating direction method of multipliers," in *Proceedings of the 31st International Conference on Machine Learning*, 2014, pp. 620–628.